PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-268034
(43)Date of publication of application : 29.09.2000

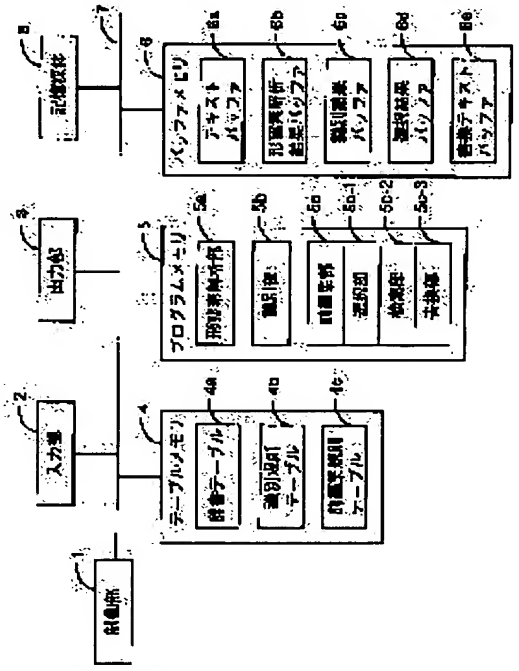
(51)Int.Cl. G06F 17/27
G06F 17/30

(21)Application number : 11-070312
(22)Date of filing : 16.03.1999

(71)Applicant : SHARP CORP
(72)Inventor : YOSHIMI TAKEHIKO

(54) AUTOMATIC TEXT PRE-EDITING DEVICE, ITS METHOD AND STORAGE MEDIUM TO BE USED FOR IT

(57)Abstract:
PROBLEM TO BE SOLVED: To reduce a burden on a user by automating the identification of a text kind and the selection of a pre-editing rule and standard notation corresponding to the kind to rewrite in the pre-editing processing of a text described in natural language.
SOLUTION: When an input part 2 inputs a text described in natural language, the morpheme analytic part 5a of a program memory 5 morpheme- analyzes each word of the inputted text by referring to a dictionary table 4a in a table memory 4, and extracts part-of-speech information and morpheme information. Next, an identification part 5b identifies the kind of the text from extracted part-of-speech information and morpheme information by referring to an identification rule table 4b. Then a pre-editing part 5c detects a word at the pre-editing object part of the text from an editing rule corresponding to the identified kind by referring to an editing rule table 4c, and writes it to standard notation. Thus, selecting processing of a group of pre-editing rules according to various kinds of text kinds is automated to reduce the burden on the user.



LEGAL STATUS

[Date of request for examination] 27.07.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

公開特許・実用（抄録Ａ）

第P2000-268034号

【名称】テキスト自動前編集装置及び方法並びにこれに利用される記憶媒体

審査／評価者請求　未　　請求項／発明の数　8　（公報　8頁、抄録　6頁）　　公開日　平成12年(2000) 9月29日

出願／権利者　シャープ株式会社（大阪府大阪市阿倍野区長池町２２番２２号）
発明／考案者　吉見　毅彦

出願番号　特願平11-70312　　平成11年(1999) 3月16日
代理人　野河　信太郎

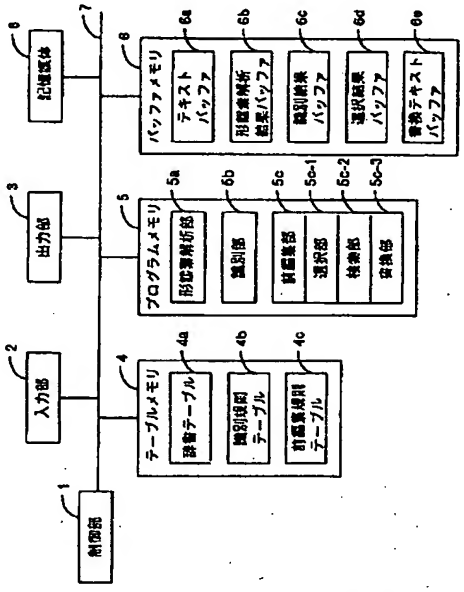
Int.Cl.7　識別記号
G06F 17/27
17/30
FI
G06F 15/38
15/40　370
15/401　320
5B075ND03;NS01;PP02;PP04;PP10;PQ02;F
5B091AA06;AA15;CA02;CB02;CB09;CB14;C

【発明の属する技術分野】本発明は、自然言語処理システムの機械翻訳装置などに適用され、自然言語で記述されたテキストをその意味を変えない範囲で前編集することによって機械翻訳などの自然言語処理精度の向上を図るテキスト前編集装置及び方法並びにこれに利用される記憶媒体に関する。

(57)【要約】

【課題】　自然言語で記述されたテキストの前編集処理において、テキストの種類の識別と、そのテキストの種類に応じた前編集規則及び標準表記の選択を自動化する。

【解決手段】　テキストに関連する単語、品詞情報、形態素情報を記憶した辞書テーブルと、テキストの種類を識別するための識別規則を記憶した識別規則テーブルと、テキストを前編集するための前編集規則及び標準表記をテキストの種類別に記憶した前編集規則テーブルと、テキストを入力する入力部と、辞書テーブルを参照し、入力テキストの各単語について形態素解析して品詞情報、形態素情報を抽出する形態素解析部と、識別規則テーブルを参照し、抽出された品詞情報、形態素情報からテキストの種類を識別する識別部と、前編集規則テーブルを参照し、識別された種類に対応する編集規則からテキストの前編集対象部分の単語を検出し、その単語を標準表記に書き換える前編集部とから構成される。



【特許請求の範囲】

【請求項１】　自然言語で記述されたテキストに関連する単語、品詞情報、形態素情報を記憶した辞書テーブルとテキストの種類を識別するための識別規則を記憶した識別規則テーブルとテキストを前編集するための前編集規則及び標準表記をテキストの種類別に記憶した前編集規則テーブルとからなるテーブルメモリと、テキストを入力する入力部と、辞書テーブルを参照し、入力テキストの各単語について形態素解析して品詞情報、形態素情報を抽出する形態素解析部と、識別規則テーブルを参照し、抽出された品詞情報、形態素情報からテキストの種類を識別する識別部と、前編集規則テーブルを参照し、識別された種類に対応する編集規則からテキストの前編集対象部分の単語を検出し、その単語を標準表記に書き換える前編集部とを備えたことを特徴とするテキスト自動前編集装置。

【請求項２】　前記前編集部は、前編集規則テーブルを参照し、前記識別部により識別されたテキストの種類に対応する編集規則を選択する選択部と、選択された編集規則に対応するテキストの前編集対象部分の単語を検索する検索部と、検索された単語を標準表記に書き換える書換部とをさらに備えたことを特徴とする請求項１記載のテキスト自動前編集装置。

【請求項３】　前記識別規則テーブルは、テキストの種類が科学技術論文、特許明細書、機器の取り扱い説明書または報道記事であるか否かを識別するための識別規則を分野別に記憶したことを特徴とする請求項１記載のテキスト自動前編集装置。

【請求項４】　自然言語で記述されたテキストに関連する単語、品詞情報、形態素情報を辞書テーブルに記憶し、テキストの種類を識別するための識別規則を識別規則テーブルに記憶し、テキストを前編集するための前編集規則及び標準表記をテキストの種類別に前編集規則テーブルを記憶し、入力部を用いて、テキストを入力し、形態素解析部を用いて、辞書テーブルを参照し、入力テキストの各単語について形態素解析して品詞情報、形態素情報を抽出し、識別部を用いて、識別規則テーブルを参照し、抽出された品詞情報、形態素情報からテキストの種類を識別し、前編集部を用いて、前編集規則テー

ブルを参照し、識別された種類に対応する編集規則からテキストの前編集対象部分の単語を検出し、その単語を標準表記に書き換えることを特徴とするテキスト自動前編集方法。

【請求項５】　コンピュータに、自然言語で記述されたテキストに関連する単語、品詞情報、形態素情報を記憶した辞書テーブルを参照させ、入力テキストを形態素解析して品詞情報、形態素情報を抽出させ、テキストの種類を識別するための識別規則を記憶した識別規則テーブルを参照させ、抽出された品詞情報、形態素情報からテキストの種類を識別させ、テキストを前編集するための前編集規則及び標準表記をテキストの種類別に記憶した前編集規則テーブルを参照し、識別された種類に対応する編集規則からテキストの前編集対象部分の単語を検出させ、その単語を標準表記に書き換えさせる処理を実行させるテキスト自動前編集プログラムを記憶した記憶媒体。

【請求項６】　異なる言語間の機械翻訳を行う機械翻訳装置が請求項１記載のテキスト自動前編集装置を用いて翻訳の前編集処理を行うことを特徴とする機械翻訳システム。

【請求項７】　自然言語装置を接続と自然言語で記述されたテキストの転送を制御するインターフェイスが請求項１記載のテキスト自動前編集装置を用いてテキストの前編集処理を行うことを特徴とする自然言語インターフェイス。

【請求項８】　自然言語で記述されたテキストを自動要約するテキスト要約装置が請求項１記載のテキスト自動前編集装置を用いてテキストを前編集処理を行うことを特徴とするテキスト要約システム。

【発明の実施の形態】本発明の構成において、（１）既存規則は変更せず、特殊な表現を扱うための規則群を、標準的な表現を扱うための既存規則群から独立させた形式で保持した前編集規則テーブルと、既存規則による処理を行う前に、既存規則でも適切に処理できるように特殊な形式の表現を標準的な表現に書き換える前編集部を設ける。（２）対象テキストに含まれる様々な種類の情報を、識別規則テーブルに基づいてテキストの種類を自動的に識別し、その識別結果に従って適切な前編集規則群を前編集規則テーブルから自動的に選択する。

なお、本発明において、辞書テーブル、識別規則テーブル、前編集規則テーブルは、例えば、本体と分離可能な磁気テープやカセットテープ等のテープ系、フロッピー（登録商標）ディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系、あるいはマスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリを含めた固定的にプログラムを担持する記憶媒体で構成してもよい。入力部は、キーボード、マウス、ペン、タブレット、スキャナ、文字認識装置記憶媒体読取装置などの入力装置で構成してもよい。識別部、前編集部は、例えば、ＣＰＵ、ＲＯＭ、ＲＡＭ、Ｉ／Ｏポートからなるコンピュータ、ＣＰＵを含むＡＳＩＣなどで構成してもよい。

前記前編集部は、前編集規則テーブルを参照し、前記識別部により識別されたテキストの種類に対応する編集規則を選択する選択部と、選択された編集規則に対応するテキストの前編集対象部分の単語を検索する検索部と、検索された単語を標準表記に書き換える書換部とをさらに備えた構成にしてもよい。この構成において、選択部、検索部、書換部は、例えば、ＣＰＵ、ＲＯＭ、ＲＡＭ、Ｉ／Ｏポートからなるコンピュータ、ＣＰＵを含むＡＳＩＣなどで構成してもよい。

前記識別規則テーブルは、テキストの種類が科学技術論文、特許明細書、機器の取り扱い説明書または報道記事であるか否かを識別するための識別規則を分野別に記憶した構成にしてもよい。

この構成によれば、異なる言語間の機械翻訳を行う機械翻訳装置が本発明のテキスト自動前編集装置を用いて翻訳の前編集処理を行う機械翻訳システムが提供される。また、自然言語装置を接続と自然言語で記述されたテキストの転送を制御するインターフェイスが本発明のテキスト自動前編集装置を用いてテキストの前編集処理を行う自然言語インターフェイスが提供される。また、自然言語で記述されたテキストを自動要約するテキスト要約装置が本発明のテキスト自動前編集装置を用いてテキストを前編集処理を行うテキスト要約システムが提供される。

以下、図に示す実施例に基づいて本発明を詳述する。なお、これによって本発明は限定されることはない。

図１は本発明の一実施例であるテキスト自動前編集装置の構成を示すブロック図である。図１において、１はコピュータのＣＰＵ（中央処理装置）からなる制御部を示し、制御部１は、プログラムメモリに記憶された制御プログラムにより各部を制御する。

２はキーボード、マウス、ペン、タブレット、スキャナ、文字認識装置などの入力装置や、通信回線と接続されている通信装置、記憶媒体読取装置などからなる入力部を示し、入力部２は自然言語で記述されたテキストの入力、前編集処理の指示、テキストの通信、制御プログラムのインストールなどを行う。

３はＣＲＴ（陰極線管）ディスプレイ、ＬＣＤ（液晶ディスプレイ）、ＰＤ（プラズマディスプレイ）などからなる表示装置３ａや、サーマルプリンタ、レーザプリンタなどからなる印字装置、または通信回線と接続されている通信装置３ｃで構成される出力部を示し、出力部３は、入力部２による入力結果、制御部１の制御により翻訳結果を表示装置３ａに表示したり、印字装置３ｂを介して印字したり、通信装置３ｃを介して送信する。

４はマスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた記憶媒体からなるテーブルメモリを示し、テーブルメモリ４は、単語、品詞情報、形態素情報を記憶した辞書テーブル４ａ、テキストの種類を識別するための識別規則を記憶した識別規則テーブル４ｂ、テキストを前編集するための新聞記事の見出し用前編集規則、特許明細書用前編集規則、取り扱い説明書用前編集規則などの前編集規則及び標準表記をテキストの種類別に記憶した前編集規則テーブル４ｃとして機能する。

５はマスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系

等を含めた記憶媒体からなるプログラムメモリを示し、プログラムメモリ５は、辞書テーブル４ａを参照し、入力テキストの各単語について形態素解析して品詞情報、形態素情報を抽出する形態素解析部５ａ、識別規則テーブル４ｂを参照し、形態素解析部５ａによって抽出された単語の品詞情報、形態素情報からテキストの種類を識別する識別部５ｂ、前編集規則テーブル４ｃを参照し、識別された種類に対応する編集規則からテキストの前編集対象部分の単語を検出し、その単語を標準表記に書き換える前編集部５ｃとして機能する各制御プログラムを記憶している。

前編集部５ｃは、識別部５ｂにより識別されたテキストの種類に対応する編集規則を前編集規則テーブルから選択する選択部５ｃ-1と、選択された編集規則に対応するテキストの前編集対象部分の単語を検索する検索部５ｃ-2と、検索された単語を標準表記に書き換える書換部５ｃ-3として機能する。

６はＲＡＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた記憶媒体からなるバッファメモリを示し、バッファメモリ６は、入力テキストを記憶するテキストバッファ６ａ、形態素解析部５ａで形態素解析された単語、品詞情報、形態素情報を記憶する形態素解析結果バッファ６ｂ、識別部５ｂで識別された種類を記憶する識別結果バッファ６ｃ、選択部５ｃ-1で選択された標準表記を記憶する選択結果バッファ６ｄ、書換部５ｃ-3で書き換えられたテキストを記憶する書換テキストバッファ６ｅとして機能する領域に備えている。

制御部１は、識別規則とのマッチング処理したデータや前編集規則とのマッチング処理したデータを各バッファに記憶する。７はバスラインを示し、制御プログラムデータ及びアドレスデータが転送される。制御部１は、バスライン７を会してプログラムメモリ４から制御プログラムを読み出して各部を制御することにより本発明のテキスト前編集装置を実現する。

８はマスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた本体と分離可能なメディアで構成した固定的にプログラムを担持する記憶媒体を示し、記憶媒体８に本発明のテキスト前編集プログラムを記憶し、入力部２の記憶媒体読取装置を介してバッファメモリ６の予備領域にテキスト前編集プログラムをインストールすることにより本発明のテキスト前編集機能を実現してもよい。また、この記憶媒体は、本テキスト前編集装置がインターネットを含めた外部の通信ネットワークとの接続可能な通信装置を備えている場合には、その通信装置を介して通信ネットワークからプログラムをダウンロードするように流動的にプログラムを担持する媒体であってもよい。なお、このように通信ネットワークからプログラムをダウンロードする場合には、そのダウンロード用プログラムは予め本体装置に格納しておくか、あるいは別な記憶媒体からインストールされるものであってもよい。なお、記憶媒体に格納されている内容としてはプログラムに限定されず、データであってもよい。

本発明の別の観点によれば、自然言語で記述されたテキストに関連する単語、品詞情報、形態素情報を記憶した辞書テーブル４ａとテキストの種類を識別するための識別規則を記憶した識別規則テーブル４ｂとテキストを前編集するための前編集規則及び標準表記をテキストの種類別に記憶した前編集規則テーブル４ｃとからなるテーブルメモリ４と、テキストを入力する入力部とを備えたテキスト自動前編集装置に用いられ、コンピュータ１によって読み取り可能なテキスト自動前編集プログラムを記憶した記憶媒体８であって、前記コンピュータ１に、辞書テーブル４ａを参照し、入力テキストを形態素解析して品詞情報、形態素情報を抽出させ、識別規則テーブル４ｂを参照し、抽出された品詞情報、形態素情報からテキストの種類を識別させ、前編集規則テーブル４ｃを参照し、識別された種類に対応する編集規則からテキストの前編集対象部分の単語を検出させ、その単語を標準表記に書き換えさせることができる。

図２は本実施例のテキスト自動前編集処理の手順を示すフローチャートである。以下、図３～図５を用いて、英日機械翻訳システムで翻訳されるテキストの前編集処理について説明する。図３は本実施例の入力テキストの一例を示す図である。図３に示すように、英日機械翻訳の対象となるＥ１～Ｅ３のテキストが入力部２により入力されテキストバッファ６ａに記憶される。続いて、入力部２により自動テキスト前編集処理の指示が入力される。

ＳＴＥＰ１では、形態素解析部５ａにより対象テキストの形態素解析を行い、対象テキストに含まれる対象表現となる各単語について品詞などの語彙属性を抽出する。ＳＴＥＰ２では、１対象テキストに含まれる対象表現の数をカウントする対象表現数カウンタの値ｉをリセットする。ＳＴＥＰ３～ＳＴＥＰ６では、識別部５ｂにより、対象テキストの先頭から一表現ずつ順に、識別規則テーブルに記憶された各テキスト識別規則とのマッチングを行い、どの条件にマッチするかに応じてテキストの種類を識別する。

図４は本実施例の識別規則テーブルに記憶された識別規則の一例を示す図である。入力テキストの種類を識別する場合、入力テキストの対象表現が、識別規則１）と２）を満たすかどうかは、対象表現を比較することによって判断する。図４では、図３に示す入力テキストのＥ２、Ｅ３は、識別規則１）と２）を満たすので、この入力テキストは新聞記事であると識別される。従来技術では、対象テキストの種類が例えば新聞記事であることをあらかじめ指定しておく必要があった。本実施例では対象テキストの種類を自動的に識別する。

ＳＴＥＰ７では、識別規則テーブルにあらかじめ記憶されている識別規則によってテキストの種類に識別できない場合、テキストの種類としてある値（例えば「一般テキスト」）が設定される。ＳＴＥＰ８では、１対象テキストに含まれる対象表現の数をカウントする対象表現数カウンタの値ｉをリセットする。

ＳＴＥＰ９～ＳＴＥＰ１１では、前編集部５ｃ（選択部、検索部、書換部）により、対象テキストの先頭から一表現ずつ順に、前編集規則群テーブルに記憶された各前編集規則とのマッチングを行い、マッチングに成功した表現に対して書き換えを行うことにより、前編集処理を実行する。なお、ＳＴＥＰ９のｎは入力テキストに

含まれる表現の数である。

図５は本実施例の前編集規則テーブルに記憶された前編集規則の一例を示す図である。図５に示すように、前編集規則テーブル４ｃには、例えば、新聞記事の見出し用の前編集規則が記憶されている。図５の例では、前編集規則１）、２）、３）がすべて満たされるとき、“ｔｏ”を“ｗｉｌｌ”に書き換える前編集処理である。

図３に示すテキストを前編集する場合、Ｅ１のテキスト“Agency to inspecthealth of 8 banks ”は、図５の前編集規則１）、２）、３）をすべて満たすので、“Agency will inspect health of 8 banks ”に書き換えられる。“Agency to inspect health of 8 banks ”という表現を“Agency will inspect health of 8 banks ”に書き換えてもよいのは、この表現が新聞記事の見出しである場合であり、そうでない場合は書き換えてよいとは限らない。

これに対して、Ｅ２、Ｅ３のテキストは、この編集規則を満たさないので書き換えられない。テキストの対象表現が、図５の前編集規則１）を満たすかどうかは、対象表現と、例えば以下のような名詞句パターンがマッチングするかどうかによって調べる。

名詞句パターン：｜冠詞｜　｜｜副詞｜限定形容詞｜｜名詞*｜名詞

ここで、括弧｜｜で囲まれた要素（冠詞、副詞、限定形容詞、名詞*）は存在してもしなくてもよいことを意味し、記号「*」は、要素が任意回数繰り返し出現可能であることを意味する。

対象表現が前編集規則２）を満たすかどうかは、対象表現上で、単語“for”を検索することによって判断する。また、対象表現が前編集規則３）を満たすかどうかは、例えば、次のような手続きによって調べる。

テキストの対象表現が前編集規則を満たすかどうかの判定手順としては、対象表現の先頭から順に、述語になり得る定形動詞を検索する。もし、述語が見つかれば、その述語候補と、人称・数が一致する名詞を主辞とする名詞句が前方に存在するかどうかを調べる。　名詞句の検索には、上記の名詞句パターンを利用する。もし、そのような名詞句が存在すれば、それを主語と見なし、対象表現全体を文と見なし、前編集規則３）が満たされないものとする。

前編集処理しないテキスト“Agency to inspect health of 8 banks ”を、標準的な表現を主な対象とした従来の機械翻訳システムに入力して翻訳すると、例えば「８つの銀行の健全性を調査するための機関」のように新聞記事の見出しとしては不適切な翻訳が生成される。

これに対して、ＳＴＥＰ１での形態素解析処理、ＳＴＥＰ２～ＳＴＥＰ７でのテキスト識別処理、ＳＴＥＰ８～ＳＴＥＰ１１でのテキスト前編集処理の説明からわかるように、本発明のテキスト前編集装置で書き換えた“Agency will inspect health of 8 banks ”を、従来の機械翻訳システムの入力して翻訳すれば、例えば、「機関は８つの銀行の健全性を調査するであろう」のように新聞記事の見出しとして適切な翻訳が得られる。

図６は本発明のテキスト自動前編集装置の適用システムの一例を示す図である。本発明のテキスト自動前編集装置は、自然言語処理システムとは独立であり、本発明のテキスト前編集装置から出力される前編集されたテキストを利用するシステムとして、図６に示すような適用テキストがある。

例えば、異なる言語間の機械翻訳を行う機械翻訳装置が本発明のテキスト自動前編集装置を用いて翻訳の前編集処理を行う機械翻訳システムを提供することができる。また、自然言語装置を接続と自然言語で記述されたテキストの転送を制御するインターフェイスが本発明のテキスト自動前編集装置を用いてテキストの前編集処理を行う自然言語インターフェイスを提供することができる。また、自然言語で記述されたテキストを自動要約するテキスト要約装置が本発明のテキスト自動前編集装置を用いてテキストを前編集処理を行うテキスト要約システムを提供することができる。

【図面の簡単な説明】

【図１】本発明の一実施例であるテキスト自動前編集装置の構成を示すブロック図である。

【図２】本実施例のテキスト自動前編集処理の手順を示すフローチャートである。

【図３】本実施例の入力テキストの一例を示す図である。

【図４】本実施例の識別規則テーブルに記憶された識別規則の一例を示す図である。

【図５】本実施例の前編集規則テーブルに記憶された前編集規則の一例を示す図である。

【図６】本発明のテキスト自動前編集装置の適用システムの一例を示す図である。

【符号の説明】

１　制御部
２　入力部
３　出力部
４　テーブルメモリ
４ａ　辞書テーブル
４ｂ　識別規則テーブル
４ｃ　前編集規則群テーブル
５　プログラムメモリ
５ａ　形態素解析部
５ｂ　識別部
５ｃ　前編集部
５ｃ-１　選択部
５ｃ-２　検索部
５ｃ-３　書換部
６　バッファメモリ
６ａ　テキストバッファ
６ｂ　形態素解析結果バッファ
６ｃ　識別結果バッファ
６ｄ　選択結果バッファ
６ｅ　書換テキストバッファ
７　バスライン
８　記憶媒体

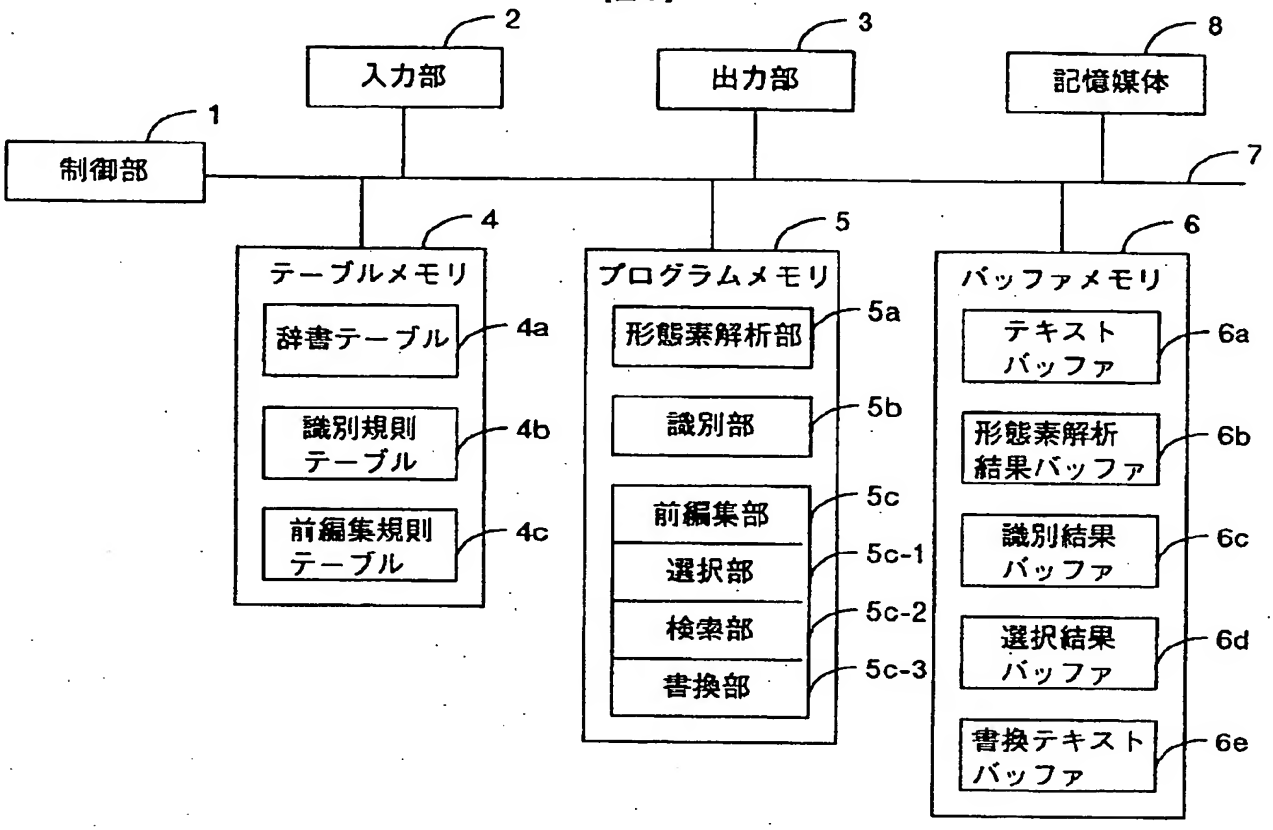
【図４】

1) テキストの冒頭付近に"By 記者名(人名)"という表現がある。
2) 記者名を含む表現の次の表現が "地名--"で始まる。
↓
対象テキストは新聞記事である。

【図５】

1) to付不定詞の直前が名詞句である。
2) to付不定詞の前方にforがない。
3) 処理対象表現が文でない。
↓
toをwillに書き換える。

【図１】

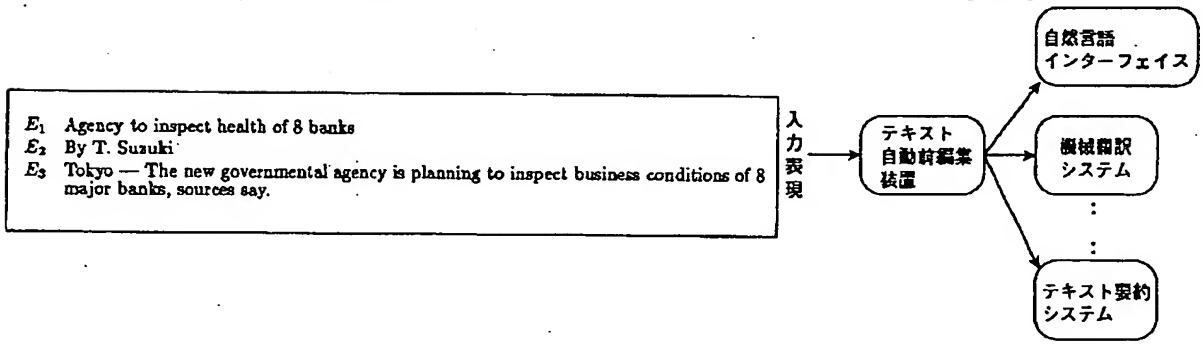


【図３】

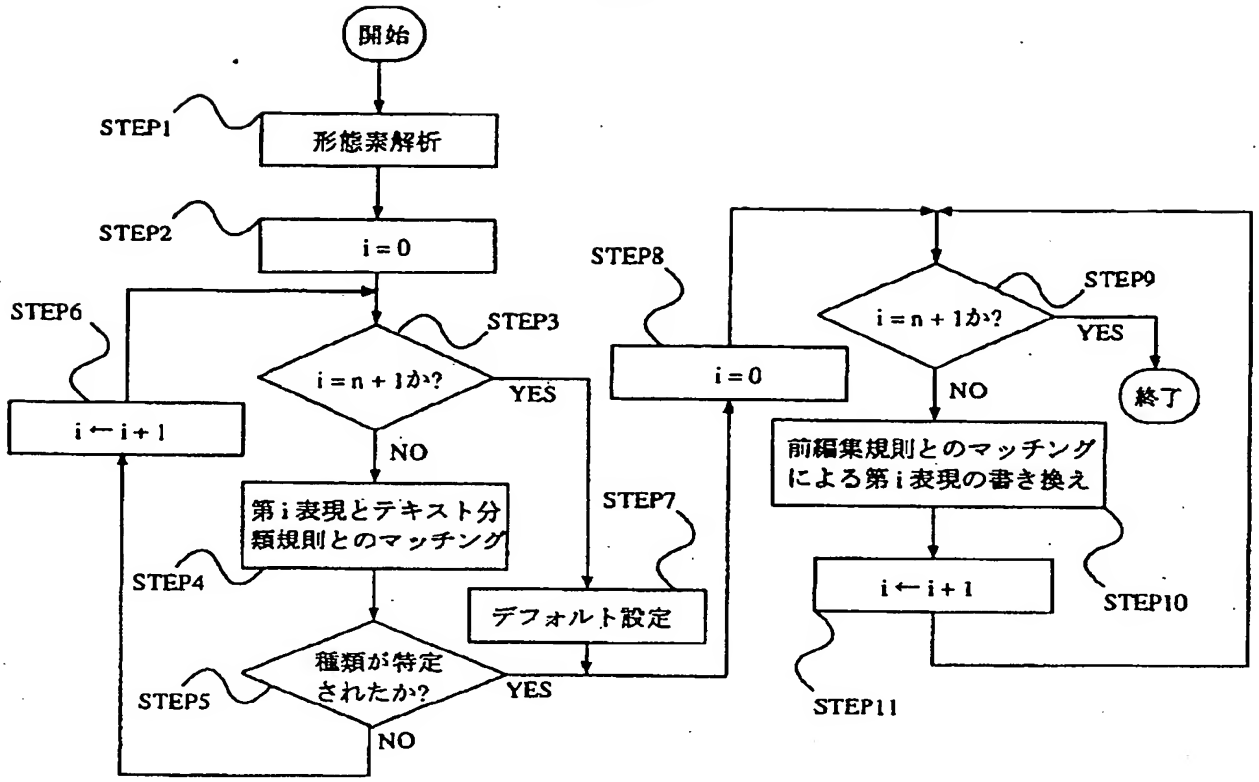
E_1 Agency to inspect health of 8 banks
E_2 By T. Suzuki
E_3 Tokyo — The new governmental agency is planning to inspect business conditions of 8 major banks, sources say.

入力表現 → テキスト自動前編集装置 → 自然言語インターフェイス / 機械翻訳システム / ： / テキスト要約システム

【図６】



【図２】



【書誌的事項の続き】
【ＦＩ】　G06F 15/38;15/40　370;15/401　320
【Ｆターム】　5B075ND03;NS01;PP02;PP04;PP10;PQ02;PQ03
5B091AA06;AA15;CA02;CB02;CB09;CB14;CB22;CB30;CC03;EA17
【識別番号または出願人コード】000005049
【出願／権利者名】　シャープ株式会社
大阪府大阪市阿倍野区長池町２２番２２号
【発明／考案者名】　吉見　毅彦
大阪府大阪市阿倍野区長池町２２番２２号　シャープ株式会社内
【代理人】　野河　信太郎(100065248)
【出願形態】OL

注）本抄録の書誌的事項は初期登録時のデータで作成されています。